

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-132976

(43)Date of publication of application : 10.05.2002

(51)Int.Cl.

G06F 17/60

G06F 13/00

G06F 17/30

(21)Application number : 2000-322631

(71)Applicant : UNIDEN CORP

(22)Date of filing : 23.10.2000

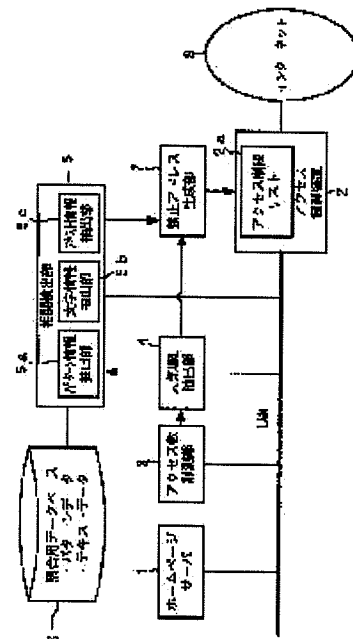
(72)Inventor : SEKI KAZUHIKO
OKAZAKI ISAO

(54) HOMEPAGE MANAGEMENT DEVICE AND METHOD, HOMEPAGE EVALUATION DEVICE AND RECORDING MEDIUM

(57)Abstract:

PROBLEM TO BE SOLVED: To automatically extract the undesirable homepages, managed by an internet service provider (ISP), to reduce the monitoring load on the ISP, and to provide the homepages without lowering service quality.

SOLUTION: This device is provided with a number-of-access measurement part for measuring the number of access for each URL(uniform resource locator) indicating the homepage, a popular URL extraction part for preparing the list of the URLs for which the predetermined number or more of the access are measured, based on the measured result of the number-of-access measurement part, a correlation detection part for collating the contents of the URLs entered in the URL list prepared by the popular URL extraction part with a database for collation and calculating a correlation value and an inhibited address generation part for defining the URL exceeding a predetermined correlation value in a correlation value list outputted by the correlation detection part as an inhibited address and adding the inhibited address to an access limit list. The undesirable URL is extracted automatically, and access to it is inhibited.



(11)特許出願公開番号

特開2002-132976

(P2002-132976A)

(43)公開日 平成14年5月10日(2002.5.10)

(51)Int.Cl. ⁷	識別記号	F I	テ-マ-ト*(参考)
G 0 6 F 17/60	1 3 2	G 0 6 F 17/60	1 3 2 5 B 0 7 5
	1 4 2		1 4 2
	1 5 0		1 5 0
	5 0 4		5 0 4
13/00	5 2 0	13/00	5 2 0 A
審査請求 未請求 請求項の数16 O L (全 11 頁) 最終頁に続く			

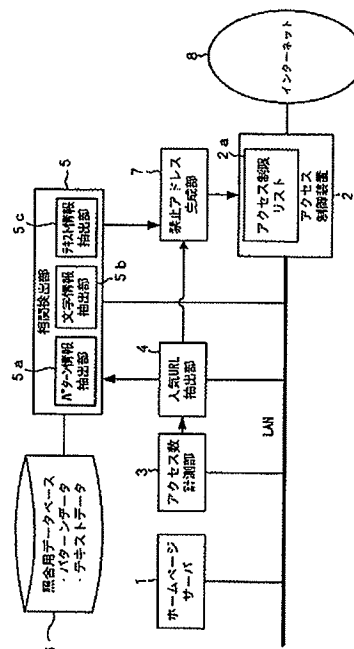
(21)出願番号	特願2000-322631(P2000-322631)	(71)出願人	000115267 ユニデン株式会社 東京都中央区八丁堀二丁目12番7号
(22)出願日	平成12年10月23日(2000.10.23)	(72)発明者	関 和彦 東京都中央区八丁堀二丁目12番7号 ユニ デン株式会社内
		(72)発明者	岡崎 功 東京都中央区八丁堀二丁目12番7号 ユニ デン株式会社内
		(74)代理人	100107113 弁理士 大木 健一 Fターム(参考) 5B075 KK54 KK63

(54)【発明の名称】 ホームページ管理装置及び方法並びにホームページ評価装置及び記録媒体

(57) 【要約】

【課題】 インターネットサービスプロバイダ（ISP）が管理しているホームページのうち、好ましくないものを自動的に抽出する。ISPの監視負担を軽減するとともに、サービス品質を下げることなくホームページを提供する。

【解決手段】 ホームページを指示するURL (Uniform Resource Locator) 毎にアクセス数を計測するアクセス数計測部と、該アクセス数計測部の計測結果に基づいて予め定められた数以上のアクセスが計測されたURLリストを作成する人気URL抽出部と、該人気URL抽出部により作成された前記URLリストに記載されたURLの内容を照合用データベースと照合して相関値を算出する相関検出部と、該相関検出部の出力する相関値リストのうち予め定められた相関値を超えているURLを禁止アドレスと定義し、該禁止アドレスを前記アクセス制限リストに追加する禁止アドレス生成部とを備える。好ましくないURLを自動的に抽出し、これに対するアクセスを禁止する。



【特許請求の範囲】

【請求項 1】 制作者が作成したホームページを登録してインターネットからの閲覧に供するホームページサーバと、該ホームページへのインターネットからのアクセスをアクセス制限リストに基づき制御するアクセス制御部とを備えるインターネット・サービス・システムを管理するためのホームページ管理装置であって、ホームページを指示する URL (Uniform Resource Locator) 毎にアクセス数を計測するアクセス数計測部と、該アクセス数計測部の計測結果に基づいて予め定められた数以上のアクセスが計測された URL リストを作成する人気 URL 抽出部と、

該人気 URL 抽出部により作成された前記 URL リストに記載された URL の内容を照合用データベースと照合して相関値を算出する相関検出部と、該相関検出部の出力する相関値リストのうち予め定められた相関値を超えている URL を禁止アドレスと定義し、該禁止アドレスを前記アクセス制限リストに追加する禁止アドレス生成部とを備えるホームページ管理装置。

【請求項 2】 前記相関検出部は、画像情報からパターン抽出を行うパターン情報抽出部を備え、前記相関検出部が前記ホームページサーバに登録されているホームページを読み出す際に、前記パターン情報抽出部を介して得られたパターン情報を前記照合用データベースに含まれるパターン・データと照合して相関値を算出することを特徴とする、請求項 1 記載のホームページ管理装置。

【請求項 3】 前記相関検出部は、画像情報から文字情報を抽出する文字情報抽出部を備え、前記相関検出部が前記ホームページサーバに登録されているホームページを読み出す際に、前記文字情報抽出部を介して得られた文字列を前記照合用データベースに含まれるテキスト・データと照合して相関値を算出することを特徴とする、請求項 1 記載のホームページ管理装置。

【請求項 4】 前記相関検出部は、テキスト・データを抽出するテキスト抽出部を備え、前記相関検出部が前記ホームページサーバに登録されているホームページを読み出す際に、前記テキスト抽出部を介して得られた文字列を前記照合用データベースに含まれるテキスト・データと照合して相関値を算出することを特徴とする、請求項 1 記載のホームページ管理装置。

【請求項 5】 前記照合用データベースは、画像を照合するための画像用パターン・データ及びテキストを照合するためのテキスト用パターン・データを含むことを特徴とする、請求項 1 記載のホームページ管理装置。

【請求項 6】 前記照合用データベースの前記テキスト用パターン・データは、演奏者名、作詞・作曲者名、曲名等の著作権に関わる情報を含むことを特徴とする、請求項 5 記載のホームページ管理装置。

【請求項 7】 前記照合用データベースの前記テキスト用パターン・データは、公序良俗に反するような情報を

含むことを特徴とする、請求項 5 記載のホームページ管理装置。

【請求項 8】 前記照合用データベースの前記画像用パターン・データは、CD ジャケット、本の表紙等の著作権に関する情報を含むことを特徴とする、請求項 5 記載のホームページ管理装置。

【請求項 9】 前記照合用データベースの前記画像用パターン・データは、ひとつ又は複数の予め定められた色彩及び当該色彩で表示される領域の大きさに関する情報を含み、

前記相関検出部は、前記 URL の内容に含まれる画像のうちで前記予め定められた色彩に一致する領域を求め、その割合を相関値として出力し、

前記禁止アドレス生成部は、前記相関値が前記画像用パターン・データに含まれる領域の大きさを超えるときに前記 URL を禁止アドレスと定義することを特徴とする、請求項 5 記載のホームページ管理装置。

【請求項 10】 画像を照合するための画像用パターン・データ及びテキストを照合するためのテキスト用パターン・データを含む照合用データベースと、ホームページを指示する URL (Uniform Resource Locator) 毎にその内容を前記照合用データベースと照合して相関値を算出し、これを URL の評価値として出力する相関検出部とを備えるホームページ評価装置。

【請求項 11】 前記相関検出部は、画像情報からパターン抽出を行うパターン情報抽出部を備え、前記相関検出部が前記ホームページを読み出す際に、前記パターン情報抽出部を介して得られたパターン情報を前記照合用データベースに含まれるパターン・データと照合して相関値を算出することを特徴とする、請求項 10 記載のホームページ評価装置。

【請求項 12】 前記相関検出部は、画像情報から文字情報を抽出する文字情報抽出部を備え、前記相関検出部が前記ホームページを読み出す際に、前記文字情報抽出部を介して得られた文字列を前記照合用データベースに含まれるテキスト・データと照合して相関値を算出することを特徴とする、請求項 10 記載のホームページ評価装置。

【請求項 13】 前記相関検出部は、テキスト・データを抽出するテキスト抽出部を備え、前記相関検出部が前記ホームページを読み出す際に、前記テキスト抽出部を介して得られた文字列を前記照合用データベースに含まれるテキスト・データと照合して相関値を算出することを特徴とする、請求項 10 記載のホームページ評価装置。

【請求項 14】 制作者が作成したホームページを登録してインターネットからの閲覧に供するホームページサーバと、該ホームページへのインターネットからのアクセスをアクセス制限リストに基づき制御するアクセス制御部と、画像を照合するための画像用パターン・データ

及びテキストを照合するためのテキスト用パターン・データを含む前記照合用データベースとを備えるインターネット・サービス・システムを管理するためのホームページ管理方法であって、

ホームページを指示するURL (Uniform Resource Locator) 毎にアクセス数を計測するアクセス数計測ステップと、

該アクセス数計測ステップの計測結果に基づいて予め定められた数以上のアクセスが計測されたURLリストを作成する人気URL抽出ステップと、

該人気URL抽出ステップにより作成された前記URLリストに記載されたURLの内容を前記照合用データベースと照合して相関値を算出する相関検出ステップと、

該相関検出ステップにより得られた相関値リストのうち予め定められた相関値を超えているURLを禁止アドレスと定義し、該禁止アドレスを前記アクセス制限リストに追加する禁止アドレス生成ステップとを備えるホームページ管理方法。

【請求項 15】 所定時間経過後に前記人気URL抽出ステップで用いる予め定められたアクセス数を減少させるしきい値変更ステップを備え、
前記しきい値変更ステップの後に前記アクセス数計測ステップ乃至前記禁止アドレス生成ステップを繰り返すことを特徴とする請求項 14 記載のホームページ管理方法。

【請求項 16】 制作者が作成したホームページを登録してインターネットからの閲覧に供するホームページサーバと、該ホームページへのインターネットからのアクセスをアクセス制限リストに基づき制御するアクセス制御部と、画像を照合するための画像用パターン・データ及びテキストを照合するためのテキスト用パターン・データを含む前記照合用データベースとを備えるインターネット・サービス・システムを管理するためのホームページ管理方法をコンピュータに実行させるためのプログラムを記録した媒体であって、
前記ホームページ管理方法は、
ホームページを指示するURL (Uniform Resource Locator) 毎にアクセス数を計測するアクセス数計測ステップと、
該アクセス数計測ステップの計測結果に基づいて予め定められた数以上のアクセスが計測されたURLリストを作成する人気URL抽出ステップと、
該人気URL抽出ステップにより作成された前記URLリストに記載されたURLの内容を前記照合用データベースと照合して相関値を算出する相関検出ステップと、
該相関検出ステップにより得られた相関値リストのうち予め定められた相関値を超えているURLを禁止アドレスと定義し、該禁止アドレスを前記アクセス制限リストに追加する禁止アドレス生成ステップとを備えることを特徴とする記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、インターネットにおいて提供されるホームページの管理装置及び方法、ホームページ評価装置及び当該方法を実施するためのプログラムを記録した媒体に関する。

【0002】

【従来の技術】インターネットの普及に伴い、通常の社会で発生するさまざまな現象や問題がインターネットの中でも発生するようになってきている。多くの場合、利用者がISP (Internet Service Provider) と契約し、そのホームページサーバを借りて自分のホームページを構築する。ISPは利用者(制作者)の委託を受け、インターネットからの閲覧に供するための設備を設置し、これの維持保守を行っている。

【0003】しかし、ISPは、表現の自由の問題と関連して、その内容までは吟味することができない。そのため、公序良俗に反するページが多数公開されている状況にある。もっとも、公開されているページを監視しようとしても、そのページ数の多さのため、それらを取り締まるために有効な手法がないという問題がある。一方、利用者の良心に依存した相互監視を行い、有害だと思われるページの公開中止を、公開を請け負っているISPに依頼しようとする問題もある。

【0004】

【発明が解決しようとする課題】この発明は、ISP自らが公開を請け負っているホームページを自律的に監視し、公開することにより、問題がある(例えば公序良俗に反する)と判断されるページを自発的に公開中止に誘導することができるホームページの管理装置及び方法、ホームページ評価装置及び当該方法を実施するためのプログラムを記録した媒体を提供することを目的とする。

【0005】

【課題を解決するための手段】係る課題を解決するために、この発明に係るホームページ管理装置は、制作者が作成したホームページを登録してインターネットからの閲覧に供するホームページサーバと、該ホームページへのインターネットからのアクセスをアクセス制限リストに基づき制御するアクセス制御部とを備えるインターネット・サービス・システムを管理するためのホームページ管理装置であって、ホームページを指示するURL (Uniform Resource Locator) 毎にアクセス数を計測するアクセス数計測部と、該アクセス数計測部の計測結果に基づいて予め定められた数以上のアクセスが計測されたURLリストを作成する人気URL抽出部と、該人気URL抽出部により作成された前記URLリストに記載されたURLの内容を照合用データベースと照合して相関値を算出する相関検出部と、該相関検出部の出力する相関値リストのうち予め定められた相関値を超えているURLを禁止アドレスと定義し、該禁止アドレスを前記アクセス制限リストに

追加する禁止アドレス生成部とを備えるものである。

【0006】好ましくは、前記相関検出部は、画像情報からパターン抽出を行うパターン情報抽出部を備え、前記相関検出部が前記ホームページサーバに登録されているホームページを読み出す際に、前記パターン情報抽出部を介して得られたパターン情報を前記照合用データベースに含まれるパターン・データと照合して相関値を算出する。

【0007】好ましくは、前記相関検出部は、画像情報から文字情報を抽出する文字情報抽出部を備え、前記相関検出部が前記ホームページサーバに登録されているホームページを読み出す際に、前記文字情報抽出部を介して得られた文字列を前記照合用データベースに含まれるテキスト・データと照合して相関値を算出する。

【0008】好ましくは、前記相関検出部は、テキスト・データを抽出するテキスト抽出部を備え、前記相関検出部が前記ホームページサーバに登録されているホームページを読み出す際に、前記テキスト抽出部を介して得られた文字列を前記照合用データベースに含まれるテキスト・データと照合して相関値を算出する。

【0009】好ましくは、前記照合用データベースは、画像を照合するための画像用パターン・データ及びテキストを照合するためのテキスト用パターン・データを含む。

【0010】好ましくは、前記照合用データベースの前記テキスト用パターン・データは、演奏者名、作詞・作曲者名、曲名等の著作権に関わる情報を含む。

【0011】好ましくは、前記照合用データベースの前記テキスト用パターン・データは、公序良俗に反するような情報を含む。

【0012】好ましくは、前記照合用データベースの前記画像用パターン・データは、CDジャケット、本の表紙等の著作権に関する情報を含む。

【0013】好ましくは、前記照合用データベースの前記画像用パターン・データは、ひとつ又は複数の予め定められた色彩及び当該色彩で表示される領域の大きさに関する情報を含み、前記相関検出部は、前記URLの内容に含まれる画像のうちで前記予め定められた色彩に一致する領域を求め、その割合を相関値として出力し、前記禁止アドレス生成部は、前記相関値が前記画像用パターン・データに含まれる領域の大きさを超えるときに前記URLを禁止アドレスと定義する。

【0014】この発明に係るホームページ評価装置は、画像を照合するための画像用パターン・データ及びテキストを照合するためのテキスト用パターン・データを含む照合用データベースと、ホームページを指示するURL (Uniform Resource Locator) 毎にその内容を前記照合用データベースと照合して相関値を算出し、これをURLの評価値として出力する相関検出部とを備えるものである。

【0015】好ましくは、前記相関検出部は、画像情報からパターン抽出を行うパターン情報抽出部を備え、前記相関検出部が前記ホームページを読み出す際に、前記パターン情報抽出部を介して得られたパターン情報を前記照合用データベースに含まれるパターン・データと照合して相関値を算出する。

【0016】好ましくは、前記相関検出部は、画像情報から文字情報を抽出する文字情報抽出部を備え、前記相関検出部が前記ホームページを読み出す際に、前記文字情報抽出部を介して得られた文字列を前記照合用データベースに含まれるテキスト・データと照合して相関値を算出する。

【0017】好ましくは、前記相関検出部は、テキスト・データを抽出するテキスト抽出部を備え、前記相関検出部が前記ホームページを読み出す際に、前記テキスト抽出部を介して得られた文字列を前記照合用データベースに含まれるテキスト・データと照合して相関値を算出する。

【0018】この発明に係るホームページ管理方法は、制作者が作成したホームページを登録してインターネットからの閲覧に供するホームページサーバと、該ホームページへのインターネットからのアクセスをアクセス制限リストに基づき制御するアクセス制御部と、画像を照合するための画像用パターン・データ及びテキストを照合するためのテキスト用パターン・データを含む前記照合用データベースとを備えるインターネット・サービス・システムを管理するためのホームページ管理方法であって、ホームページを指示するURL (Uniform Resource Locator) 毎にアクセス数を計測するアクセス数計測ステップと、該アクセス数計測ステップの計測結果に基づいて予め定められた数以上のアクセスが計測されたURLリストを作成する人気URL抽出ステップと、該人気URL抽出ステップにより作成された前記URLリストに記載されたURLの内容を前記照合用データベースと照合して相関値を算出する相関検出ステップと、該相関検出ステップにより得られた相関値リストのうち予め定められた相関値を超えているURLを禁止アドレスと定義し、該禁止アドレスを前記アクセス制限リストに追加する禁止アドレス生成ステップとを備えるものである。

【0019】好ましくは、所定時間経過後に前記人気URL抽出ステップで用いる予め定められたアクセス数を減少させるしきい値変更ステップを備え、前記しきい値変更ステップの後に前記アクセス数計測ステップ乃至前記禁止アドレス生成ステップを繰り返す。

【0020】この発明に係る記録媒体は、制作者が作成したホームページを登録してインターネットからの閲覧に供するホームページサーバと、該ホームページへのインターネットからのアクセスをアクセス制限リストに基づき制御するアクセス制御部と、画像を照合するための画像用パターン・データ及びテキストを照合するための

テキスト用パターン・データを含む前記照合用データベースとを備えるインターネット・サービス・システムを管理するためのホームページ管理方法をコンピュータに実行させるためのプログラムを記録した媒体であって、前記ホームページ管理方法は、ホームページを指示するURL (Uniform Resource Locator) 毎にアクセス数を計測するアクセス数計測ステップと、該アクセス数計測ステップの計測結果に基づいて予め定められた数以上のアクセスが計測されたURLリストを作成する人気URL抽出ステップと、該人気URL抽出ステップにより作成された前記URLリストに記載されたURLの内容を前記照合用データベースと照合して相関値を算出する相関検出ステップと、該相関検出ステップにより得られた相関値リストのうち予め定められた相関値を超えているURLを禁止アドレスと定義し、該禁止アドレスを前記アクセス制限リストに追加する禁止アドレス生成ステップとを備えることを特徴とする。

【0021】媒体には、例えば、フロッピー（登録商標）ディスク、ハードディスク、磁気テープ、光磁気ディスク、CD（CD-ROM、Video-CDを含む）、DVD（DVD-Video、DVD-ROM、DVD-RAMを含む）、ROMカートリッジ、バッテリバックアップ付きのRAMメモリカートリッジ、フラッシュメモリカートリッジ、不揮発性RAMカートリッジ等を含む。

【0022】また、電話回線等の有線通信媒体、マイクロ波回線等の無線通信媒体等の通信媒体を含む。インターネットもここでいう通信媒体に含まれる。

【0023】媒体とは、何等かの物理的手段により情報（主にデジタルデータ、プログラム）が記録されているものであって、コンピュータ、専用プロセッサ等の処理装置に所定の機能を行わせることができるものである。要するに、何等かの手段でもってコンピュータにプログラムをダウンロードし、所定の機能を実行させるものであればよい。

【0024】

【発明の実施の形態】この発明の実施の形態に係る装置及び方法について図面を参照して説明する。図1は、この発明の実施の形態に係るホームページ管理装置のブロック図である。この装置は、インターネットサービスプロバイダ (ISP) の設備に組み込まれている。ISPは、ホームページサーバ1を提供することにより各制作者のホームページをインターネット8上で閲覧できるようにしている。ホームページサーバ1とインターネット8の間にはアクセス制御装置2が設けられ、インターネットの利用者からのリクエストをサーバ1に伝えたり、サーバ1のレスポンスをインターネット8に伝えたりする。アクセス制御装置2は、アクセス制限リスト（メモリ）2aを備え、このリストに基づきインターネットを通じたリクエストを制限する。例えば、アクセス制限リスト2

aにホームページサーバ1の多数のホームページのうち、これらの一部のホームページのURLが含まれているとする。これらのURLに対してアクセスがあったとき、アクセス制御装置2は当該アクセスを拒絶したり、アクセスの回数を制限したりする。ISPの管理者はアクセス制限リスト2aを管理することにより、ホームページサーバ1が提供するホームページのうちで公開に適さないホームページへのアクセスを制限することができる。

10 【0025】符号3は、ホームページサーバ1内のホームページを指示するURL (Uniform Resource Locator) 毎にアクセス数を計測するアクセス数計測部である。符号4は、該アクセス数計測部の計測結果に基づいて予め定められた数以上のアクセスが計測されたURLリストを作成する人気URL抽出部である。

20 【0026】符号5は、該人気URL抽出部により作成された前記URLリストに記載されたURLの内容を照合用データベースと照合して相関値を算出する相関検出部である。相関検出部5は、ホームページに含まれる画像に表示されるパターン情報を抽出するパターン情報抽出部5a、前記画像に表示される文字情報を抽出する文字情報抽出部5b、ホームページに含まれるテキスト情報を抽出するテキスト情報抽出部5cを備える。

【0027】これら抽出部5a～5cの出力は照合用データベース6の内容と照合され、相関値が求められる。この相関値は、抽出された内容（データ）がデータベース6の内容とどの程度一致しているかを示すものである。

30 【0028】例えば、パターンに関してひとつあるいは複数の色彩の配色と各色彩の面積の比に関して比較を行い、配色と面積比が一致すれば相関値は高く、いずれか一方が一致すれば相関値は中程度であり、両方の点で一致しなければ相関値は低い。一例をあげると、人間の裸体の画像では、肌色の領域が最も大きく、かつ、毛髪の色（黒、金など）の領域が必ず含まれる。これら以外の色彩（空の青、森林の緑など）は背景と考えられ、この領域は肌色の領域よりも小さい。このような観点からパターンデータは作成される。他の例として、商標、商号、シンボルマーク、アニメのキャラクターをパターンデータとしてもち、このパターンデータと抽出されたパターンが比較される。パターン同士の相関値の求め方として公知の手法（例えば、相互相関関数を用いる手法）が適用できる。

【0029】他方、文字については、抽出された文字（テキスト）がデータベースのテキストと比較される。一致している文字数が多いほど相関値は高い。例えば、抽出された文字とデータベースのテキストが完全一致すれば相関値は100%とし、例えば10文字のうち5文字が一致すれば相関値は50%とする。

50 【0030】なお、文字情報抽出部5bが画像から文字

を抽出する処理として、公知の文字認識手法を適用できる。

【0031】符号7は、該相関検出部の出力する相関値リストのうち予め定められた相関値を超えているURLを禁止アドレスと定義し、該禁止アドレスを前記アクセス制限リストに追加する禁止アドレス生成部である。

【0032】照合用データベース6の例を図5に示す。照合用データベース6は、画像を照合するための画像用パターン・データ及びテキストを照合するためのテキスト用パターン・データを含む。前記テキスト用パターン・データは、演奏者名、作詞・作曲者名、曲名等の著作権に関わる情報を含み、前記テキスト用パターン・データは、公序良俗に反するような情報を含み、前記画像用パターン・データは、CDジャケット、本の表紙等の著作権に関する情報を含む。これにより、相関検出部5は、ホームページ制作者が他人の著作物を使用していることや、公序良俗に反する単語を使用していることを検出できる。相関検出部5及び照合用データベース6により、ホームページの評価を行うことができる。

【0033】図2は、この発明の実施の形態の装置の処理手順を示すフローチャートである。URLごとにアクセス数を計測する(S1)。アクセス数がしきい値を超えるかどうか判定し(S2)、超える場合には当該URLを人気URLリストに追加する(S3)。S1～S3により、全てのURLのうちから一部のものを抽出するのは、装置の監視負担を軽減するためである。ISPが管理するサーバ内のホームページは莫大な数に及び、これら全てを監視することは装置の処理能力の点から実用的でない。そこで一部のURLを抽出し、抽出されたURLを監視対象としている。抽出の基準として、ここではアクセス数、つまり人気を用いている。ガー、あるホームページで問題のある内容を公開したとしても、アクセス数が少なければその影響は限定的である。他方、アクセス数が多いホームページの場合、その影響を無視できない。このような観点からアクセス数を基準にURLを抽出することが望ましい。他にも、アクセス数の監視は技術的に容易であるという理由もある。なお、急激にアクセス数が増加するような場合、何か特別の事象が発生している可能性があるから、アクセス数の増加率も監視するようにしてもよい。アクセス数の増加率が著しい場合、重点的に監視するようにしてもよい。S1～S3の処理を全てのURLについて実行する(S4、S5)。

【0034】次に、抽出されたURL(人気URL)について評価を行う。人気URLリストの内容を照合用データベースと照合して相関値を算出する(S6)。この詳しい処理は後述する。相関値がしきい値を超えるかどうか判定し(S7)、超える場合には当該URLを禁止アドレスと定義してアクセス制限リストに追加する(S8)。S6～S8の処理を全てのURLについて実行す

る(S9、S10)。

【0035】図3は相関値算出の手順のフローチャートである。対象となるホームページを読み出し(S11)、これに含まれる内容であるテキストと画像を分離する(S12)。テキストデータは直接テキスト用パターン・データとの照合が行われ、相関値が計算される(S16)。画像データは画像として再生され(S13)、そこから画像に含まれる文字情報が抽出される(S14)。S14の処理として公知の文字認識アルゴリズムを用いることができる。抽出された文字情報はテキスト用パターン・データとの照合が行われ、相関値が計算される(S16)。

【0036】画像データからは、また、画像に含まれるパターンが抽出される(S15)。S15の処理として公知のパターン抽出アルゴリズムを用いることができる。例えば、画像に商標・商号・シンボルマーク・アニメのキャラクタが含まれているとき、S15でこれらが抽出される。抽出されたマークなどが画像用パターン・データと照合される(S17)。もし、これらが一致すれば、当該ホームページで商標・商号・シンボルマーク・キャラクタが使用されている可能性が高い。もし、無断で使用されているならば商標法、著作権法に違反するおそれがあり、ISPの管理者は当該ホームページに注意を払うべきである。例えば、画像用パターン・データとして、商標、商号、シンボルマーク、アニメのキャラクタをもち、このパターンデータと抽出されたパターンが比較される。パターン同士の相関値の求め方として公知の手法(例えば、相互相関関数を用いる手法)が適用できる。

【0037】S17における画像パターン・データの照合手法として、前記照合用データベースの前記画像用パターン・データにひとつ又は複数の予め定められた色彩及び当該色彩で表示される領域の大きさに関する情報をもたせ、URLの内容に含まれる画像のうちで前記予め定められた色彩に一致する領域を求め、その割合を相関値として出力するようにしてもよい。例えば、パターンに関してひとつあるいは複数の色彩の配色と各色彩の面積の比に関して比較を行い、配色と面積比が一致すれば相関値は高く、いずれか一方が一致すれば相関値は中程度であり、両方の点で一致しなければ相関値は低い。一例をあげると、人間の裸体の画像では、肌色の領域が最も大きく、かつ、毛髪の色(黒、金など)の領域が必ず含まれる。これら以外の色彩(空の青、森林の緑など)は背景と考えられ、この領域は肌色の領域よりも小さい。

【0038】図2及び図3の処理を何度か繰り返してもよい。この場合のフローチャートを図4に示す。アクセス数のしきい値を設定し(S20)、相関値のしきい値を設定し(S21)、図2及び図3の手順によりホームページを評価し(S22)、問題のあるホームページを抽出し(S23)、抽出されたホームページを監視対象

リストに追加する（S24）。図2においては、抽出されたURLを禁止アドレスとして直ちにアクセス制限リスト2aに加えたが、図4においては監視対象リストに追加する。抽出されたURLに対するアクセスを直ちに禁止すれば外部のインターネット利用者に対して問題のある内容を閲覧させないという点で望ましいが、他方、サーバ1の契約者の権利を損なうことになる。当該URLの制作者から苦情がくることも考えられる。また、本発明の実施の形態に係る装置・方法により抽出されたからといって、当該URLが本当に有害であると断定できない。そこで、当座は抽出されたURLを監視対象リストに追加するとともに、当該URLを管理者自身が継続して監視するとともに、当該URLに対する一般利用者の意見を参照にしながら、ISPの管理者が最終的な結論を出すことが望ましい。この結論を待って当該URLをアクセス制限リストに加えるようにしてもよい。

【0039】このような観点から、図4においては、S24の後に、問題のあるホームページに対する修正・削除・廃止の勧告を行う処理（S26）、その後の監視（S27）を設けている。S26の通知は電子メールで行うことができる。S27の監視は、管理者自身が監視するとともに、本発明の実施の形態の装置・方法により監視することもできる。

【0040】一方、S24の処理の後、アクセス数のしきい値を下げて（S25）、S20～S24の処理を繰り返すようにしてもよい。これはホームページは随時変更されるために継続的な監視が必要であるという理由のほかに、監視負担を軽減するという理由もある。後者についてさらに説明すると、多数のホームページを提供しているサーバについて、本発明の実施の形態の装置・方法を最初に適用すると、場合によっては非常に多数の問題あるホームページが抽出されるかもしれない。これでは管理者の監視負担は過大なものとなる。そこで、初めのうちは極めて人気の高いホームページのみを対象とする。このように対象を限定したとしても効果的であることは前述のとおりである。この処理を繰り返して人気の高いホームページに問題がなくなれば、逐次、しきい値を下げて対象を拡大する。そして前述の処理を繰り返す。したがって、図4の処理によれば、管理者の監視負担を抑制しつつ、監視対象を徐々に広げ、最終的に全てのホームページを評価することができるのである。

【0041】なお、URLの評価結果をアクセス制限リストに加えるばかりでなく、このアクセス制限リストを外部に公表するようにしてもよい。一般利用者の立場か

ら抽出されたホームページの評価してもらい、ISPはその結果を反映して管理することができる。リストを公表することにより、当該ホームページの制作者が自発的に内容を適正なものに修正することも期待できる。

【0042】本発明は、以上の実施の形態に限定されることなく、特許請求の範囲に記載された発明の範囲内で、種々の変更が可能であり、それらも本発明の範囲内に包含されるものであることは言うまでもない。

【0043】また、本明細書において、手段とは必ずしも物理的手段を意味するものではなく、各手段の機能が、ソフトウェアによって実現される場合も包含する。さらに、一つの手段の機能が、二つ以上の物理的手段により実現されても、若しくは、二つ以上の手段の機能が、一つの物理的手段により実現されてもよい。

【0044】

【発明の効果】この発明により、ISPが管理しているホームページのうち、公開により公序良俗に反するページを自発的かつ自発的に抽出して、公開中止の措置をとることができ、ISPのサービス品質を下げることなく、ホームページを提供することができる。

【図面の簡単な説明】

【図1】 この発明の実施の形態に係るホームページ管理装置のブロック図である。

【図2】 この発明の実施の形態に係るホームページ管理装置の動作フローチャートである。

【図3】 この発明の実施の形態に係るホームページ管理装置の照合フローチャートである。

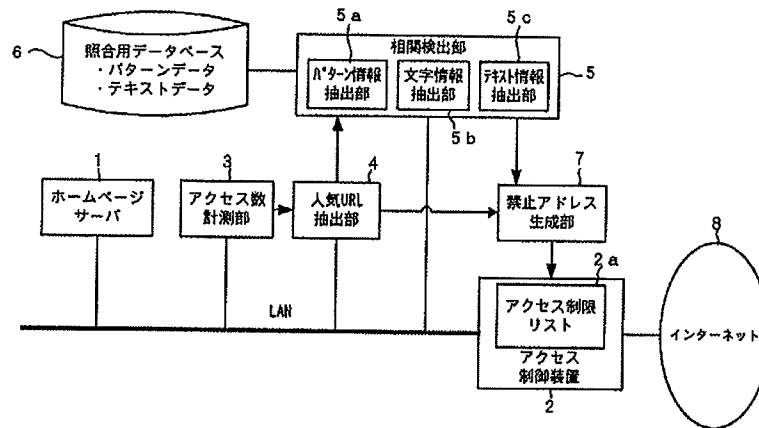
【図4】 この発明の実施の形態に係るホームページ管理装置の他の動作フローチャートである。

【図5】 この発明の実施の形態に係るホームページ管理装置の照合用データベースの説明図である。

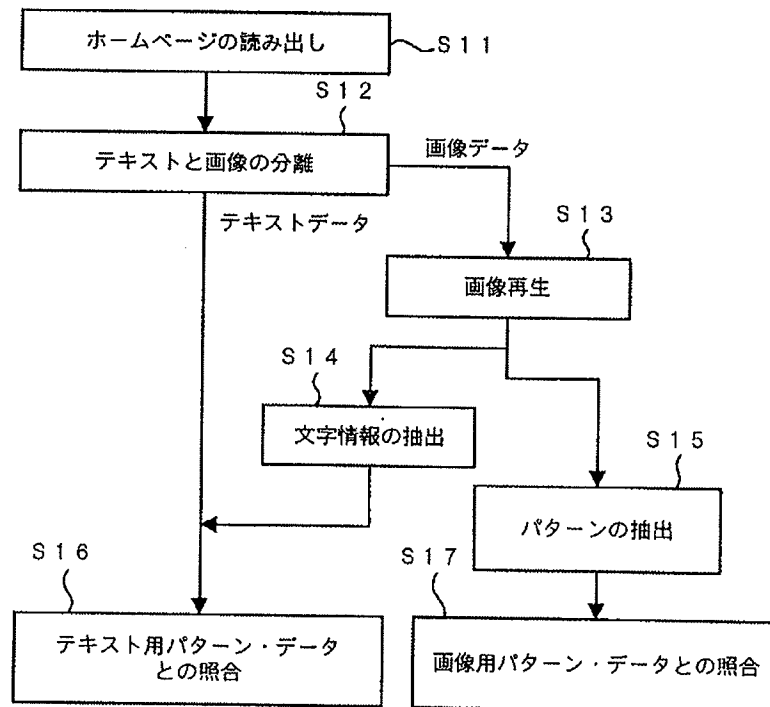
【符号の説明】

- 1 ホームページサーバ
- 2 アクセス制御装置
- 2a アクセス制限リスト
- 3 アクセス数計測部
- 4 人気URL抽出部
- 5 相関検出部
- 5a パターン情報抽出部
- 5b 文字情報抽出部
- 5c テキスト情報抽出部
- 6 照合用データベース
- 7 禁止アドレス生成部
- 8 インターネット

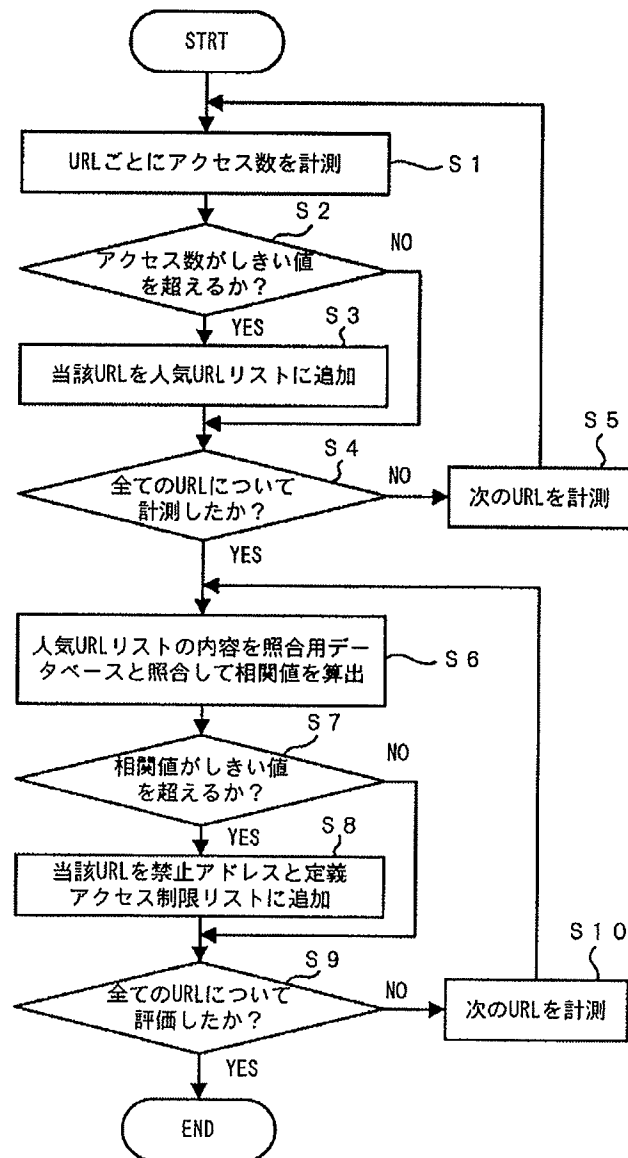
【図1】



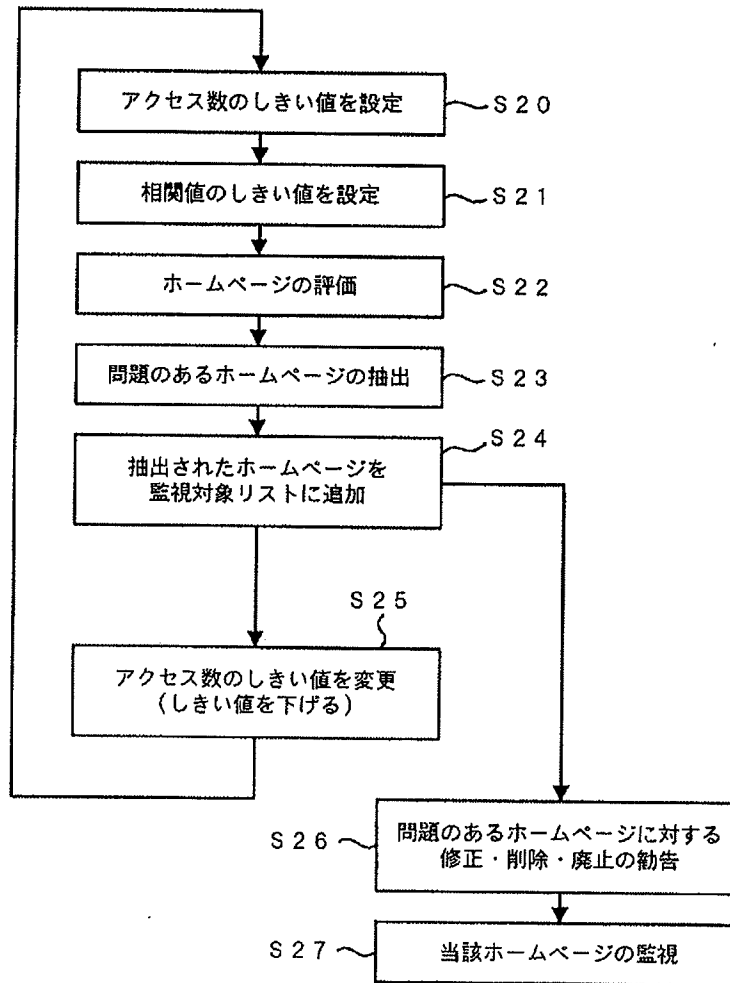
【図3】



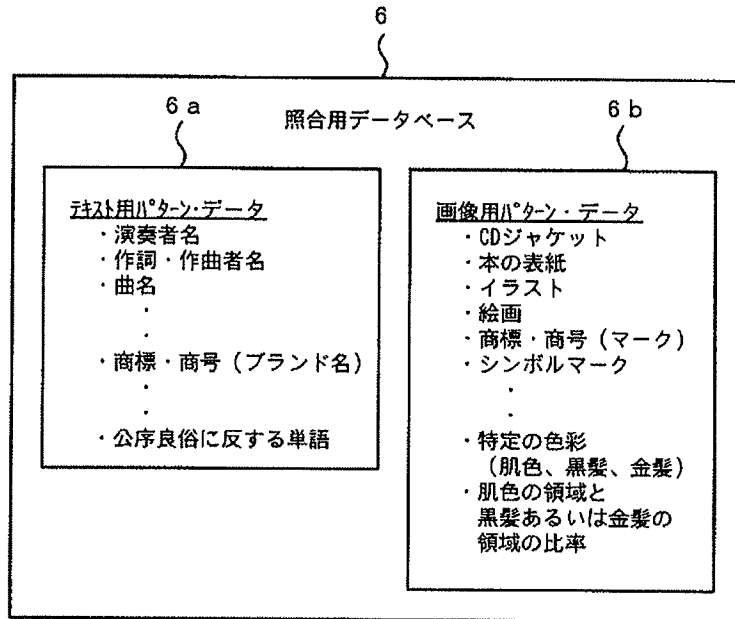
【図2】



【図 4】



【図5】



フロントページの続き

(51) Int. Cl. ⁷	識別記号	F I	ターマート* (参考)
G 0 6 F 17/30	1 1 0	G 0 6 F 17/30	1 1 0 F
	1 2 0		1 2 0 B

